An applied study on computerized adaptive testing

Schoonman, Wouter

Summary

In this thesis a study on computerized adaptive testing is reported. The goal of adaptive testing is to avoid the confrontation of testees with items that are either too easy or too difficult, and consequently, less informative for the administrator. This goal is reached by tailoring the test, during administration, to the person being tested by presenting items which have appropriate levels of difficulties given the estimated position thus far on the trait being measured. Adaptive testing can be realized by using a computer system. With a (personal) computer, item response times can be registered easily and can be used as a measure of individual differences. In the study attention was paid to this application of response time .

The study was carried out in seven modules. After a general preface and an introduction, the *construction* of four computerized parallel versions of tests in the General Aptitude Test Battery is described in chapter 2.1. In addition, the first empirical results are reported, which give evidence of a successful construction. Especially the reliability of item response times found, is remarkable high. A computerized personality questionnaire was compared with the paper and pencil version, which revealed no large differences in mean and variance of the scores. In chapter 2.2 the relation between response times and two *personality* traits (impulsivity and cognitive interference) was studied. Based on a larger sample of testees the conclusion was drawn that the possible relationships do not appear to be strong. In chapter 2.3 a limited survey of the literature on *response time* is presented. The ways in which response time could be used on a pretest to predict results on a subsequent test was studied. Response time when measured (and corrected for guessing and/or carelessness) on a simple pretest, shows correlations with total scores on subsequent tests. In the fourth chapter (2.4) the choice for *adaptive algorithms* is motivated, as well as the way in which the initial prior probability distribution can be set up. The choice was made for existing (Bayesian) algorithms, using expected a posteriori estimates for ability. The initial prior distribution was set up as uniform or as a simple combination of uniform blocks. In the following chapter (2.5) some *simulations* of adaptive testing are described. The influence of several parameters within adaptive testing has been studied. Among them are the size of the correlation between a pretest and an adaptive test, the form of the initial prior distribution, the influence of the stopping rule and the validity of the algorithms. The stopping rule had the largest influence on an adaptive testing session. The validity of the algorithms appeared to be good. In the sixth chapter (2.6), a choice was made for one of the models within item response theory, the Rasch model. An analysis was carried out on the model fit of three empirical item banks. Global statistics on model fit were computed, and a more detailed analysis was carried out for one of the banks. The conclusion was that the model seems to be violated with regard to the assumption of no guessing, especially in the case of subjects holding lower positions on the latent trait. From a practical point of view, the detected violations did not seem to make the model useless,

however. In this chapter some attention is also paid to the way in which response times, within an adaptive test, might be useful for psychodiagnostic assessment in future. The last chapter (2.7), is about an experiment carried out in practice. Over 300 applicants of Dutch Railways completed, on a voluntary basis, a pretest and an adaptive verbal test. A global estimate of the number of items generally administered in the paper based version of the Vocabulary tests resulted in an average of 32 items. In the adaptive version the average number of items is 23. The average test duration of an adaptive Vocabulary test is somewhat less than the time limit of the paper and pencil version. The attitudes of testees towards computerized testing (in comparison with paper and pencil based tests) were measured. In general, testees seemed to prefer computerized testing above paper based tests and had positive attitudes towards several aspects of computerized testing. Our adaptive test was viewed by the testees as difficult, however.

In the third part of the study conclusions were drawn on computerized testing from a psychometrical, from a technological and from an economical point of view (chapter 3.1). In chapter 3.2, topics related to adaptive testing are described. Among them are item calibration in incomplete designs, the effects of relaxing the stopping criterion, and the information that should be given to testees. The next two chapters (3.3 and 3.4) are about response time, item response theory and the way in which computerized (adaptive) testing could be realized in practice.

The fourth part of the book contains three appendices, references, an author and a subject index, a list of tables and figures and a glossary.